DOCUMENT RESUME

ED 124 568 TM 005 319

AUTHOR Melaragno, Ralph J.
TITLE Measuring Affective Development in a Longitudinal Study.
PUB DATE [Apr 76]
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Affective Behavior; *Affective Tests; Attitude Tests; Behavior Change; *Compensatory Education Programs; *Criteria; Disadvantaged Youth; Elementary Secondary Education; Locus of Control; *Longitudinal Studies; Mathematics; Motivation; Program Effectiveness; Program Evaluation; Reading; Research Problems; School Attitudes; Self Concept; Student Attitudes; Student Evaluation; Test Reviews; *Test Selection
IDENTIFIERS Elementary Secondary Education Act Title I; ESEA Title I; Survey of School Attitudes; *Sustaining Effects Study

ABSTRACT

Affective development is considered an important outcome of compensatory programs. This study examines interactions among program, cognitive, and affective variables over an extended time period. Affective dimensions considered for measurement include attitudes toward school, self-concept, achievement motivation, and internal-external control. Selecting areas for measurement is a critical task. Alternative measurement procedures are standardized instruments, available research instruments, or specially produced instruments. Each has inherent problems in a longitudinal study. Matters of special concern include linearity of affective change, comparability of different instruments, trianing of inexperienced examiners, and positive bias of very young students. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

AERA Session 13.09

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MEASURING AFFECTIVE DEVELOPMENT IN A LONGITUDINAL STUDY*

Ralph J. Melaragno

System Development Corporation

ED124568

I. RATIONALE FOR MEASURING AFFECTIVE DEVELOPMENT

The primary thrust of all compensatory education programs is to improve the cognitive development of participating students. A parallel, but subordinate, objective is the improvement of students' school-related non-cognitive development. For example, the California guidelines for Title I state:

> The main goal of every Title I project shall be to increase the academic achievement level of all eligible project participants to reflect a normal range and distribution of academic achievement for the target population as compared with the general population. Enhancing pupil self-image, motivating the pupil to achieve, improving his health, and raising his aspiration levels are to be considered supportive objectives that must be attained in order to meet the project goal.

Thus, while cognitive growth is the basic goal of a Title I project, non-cognitive improvement is viewed as an intervening variable--a precondition that enhances the likelihood of achieving cognitive growth, and that is influenced in turn by success or failure of achievement in the cognitive area.

School personnel involved in compensatory programs typically express great interest in the area that is broadly termed "attitude toward school" and that includes a student's feelings about peers, teachers, instruction, and learning. Most compensatory education program descriptions include a concern with the improvement in students' liking for different aspects of school work, particularly liking for activities in reading and mathematics (which are the main areas of concern in compensatory education).

TM005 319

*Paper presented at the AERA Convention, San Francisco, Ca., April 19-23, 1976

As part of the Sustaining Effects Study, SDC will measure students' affective development. The longitudinal nature of the study provides an opportunity to measure changes in affective behavior that occur as a student experiences the cumulative effects of a compensatory education program. SDC will be concerned with measuring the extent to which a student's affective development is influenced by participation in a compensatory education program, and the differential effects on affective behavior of programs with identifiably different characteristics. While direct assessment of affective development would be preferred (e.g., through observation of students), this is precluded by the size of the sample, the burden that would be imposed on students, and funding constraints. Thus, the measurement task must be accomplished less directly, through self-report instruments completed by students in the study.

II. METHODS OF MEASURING AFFECTIVE DEVELOPMENT

There are two methods of obtaining instruments for measuring affective behavior. One is to use existing instruments, and the second is to develop new instruments, either by creating new items or by selecting items from other scales. Each method has been used in earlier studies of compensatory education programs, and each method is accompanied by serious problems.

In its evaluation of Follow Through, Stanford Research Institute has used a variety of existing instruments. Generally, the results have pointed to limitations in the discriminating power of the instruments and a lack of strong relationships with program characteristics.

Both the Equality of Educational Opportunity study and the Emergency School Aid Act study employed the technique of embedding affective items within a questionnaire completed by students. In both cases, the scales formed by the items were not successful in distinguishing between programs with different characteristics.

In a more narrowly focused study, Educational Testing Service evaluated Compensatory Reading Programs and, having found none of the available instruments adequate to the study, developed inventories to measure students' attitudes toward reading. Early results indicate that these inventories are capable of distinguishing among different reading programs.

The major problem with using existing instruments has been that they may not be appropriate to the needs of a given study. Typically such instruments were developed for particular respondents and specific purposes, and their utility for other kinds of respondents is likely to be limited. The major problem in developing new instruments for a particular evaluation effort is the time and expense involved in determining the reliability, validity, and other psychometric properties of test items.

III. AREAS OF INTEREST AND CRITERIA FOR SELECTING INSTRUMENTS

In investigating the feasibility of using existing instruments or developing new ones in the Sustaining Effects Study, SDC first identified areas within the affective domain that appeared to be relevant to the evaluation study. Investigators looked for areas of concern to an evaluation of compensatory education; areas that have been measured in previous large-scale evaluation studies; and areas that current research has shown to be related to students' experiences in school. Application of these criteria resulted in the selection of "attitude toward school," "self-concept," "locus of control," and "achievement motivation" as the areas with the greatest relevance to the study.

Next, eight criteria for evaluating existing instruments were developed and applied to all instruments that had potential for use in the study. In an initial screening of possible instruments, standard references on affective measures and the files at the Center for the Study of Evaluation at UCLA were examined, and the eight criteria were applied to all instruments. From an

original set of over 60 instruments, 12 were judged to be of sufficient merit to warrant further consideration by SDC, with advice from a panel of experts.

The eight criteria were:

1. Validity and reliability. There should be information available (in a manual or in research literature) indicating that the instrument has acceptable construct validity and reliability.

2. Interpretability. Scores generated by the instrument should be easy to interpret for their underlying affective dimensions and should not require complicated or awkward interpretations.

3. Age appropriateness. The instrument should be valid for some or all of the ages of students in the study.

4. Administrative ease. Relatively naive examiners (in the case of this study, public school teachers) should be able to administer the instrument after limited training.

5. Scoring ease. Because of the large number of students to be measured, an instrument should be designed for, or lend itself to, machine scoring procedures.

6. Brief testing time. Because the amount of time available for measuring affective behavior is limited in the study, an instrument should not require extensive time to administer.

7. Minimal response bias. Younger students often demonstrate a bias toward socially desirable responses, and instruments should be designed to minimize this bias. This can be accomplished by both the manner in which items are prepared and the type of response called for.

8. Commonality across grades. Because of the longitudinal nature of the study, either the same or highly related instruments should be used. In particular, instruments prepared with parallel versions to encompass different grade levels were preferred.

As noted earlier, 12 instruments survived the initial screening. These instruments are described here briefly.

Animal Crackers (Adkins and Ballif). In its development form, known as Gumpgookies. Measures achievement motivation. Developed for preschool and primary-grade students. Requires 30-45 minutes testing time.

M-Scale (Williams). Measures achievement motivation. Developed for upper-grade students. Doubtful validity. Requires about 10 minutes.

Self-Esteem Inventory (Coopersmith). Measures attitude toward self in several domains. Developed for upper-grade students; has been used in primary grades in national studies. Requires about 10 minutes.

Self-Concept of Ability (Brookover). Measures academic self-concept. Developed at secondary-school level, modified for use at primary level. Requires about 20 minutes.

Children's Self-Concept Scale (Piers-Harris). Measures concerns children have about themselves. Developed for upper-grade students. Requires 15-20 minutes.

Intellectual Achievement Responsibility Questionnaire (Crandall et al.). Measures control over, and responsibility for, intellectual-academic success and failure. Developed for upper-grade students; has been used at primary grades in national studies. Requires 15-20 minutes.

Locus of Control Scale for Children (Nowicki and Strickland). Measures generalized expectancies for internal versus external control of reinforcement. Developed for middle and upper grades. Requires about 15 minutes.

Children's Locus of Control Scale (Bialer-Cromwell). Measures generalized locus of control. Developed at all elementary-grade levels. Doubtful validity. Requires about 15 minutes.

Attitude Toward Learning (Roshal et al.). Measures general attitude toward learning at school. Developed at upper elementary level. Complicated response format. Requires about 25 minutes.

Attitude Toward School (Roshal et al.). Measures attitude toward school in general. Developed at upper elementary level. Complicated response format. Requires about 25 minutes.

Attitude Toward Reading (ETS). Measures attitude toward reading instruction and reading-related activities. Forms for primary level and upper-grade level. May induce positive response bias. Requires 20-30 minutes.

Quality of School Life (Epstein and McPartland). Measures satisfaction with school, commitment to classwork, reactions to teachers. Developed for upper-grade students. Requires about 20 minutes for full scale.

In the table that follows, the 12 instruments that survived the original screening are shown, along with an indication of how each instrument fared with the eight criteria described earlier. An "X" indicates that the instrument was judged adequate on that criterion, "?" means that there was some doubt about the instrument for that criterion, while a blank indicates serious concern.

CRITERIA

INSTRUMENTS	Validity and reliability	Interpretability	Age appropriateness	Administrative ease	Scoring ease	Brief testing time	Minimal response bias	Commonality across grades
Animal Crackers	X	X	X				X	
M-Scale	?	X	X	X	X	X	X	
Self-Esteem Inventory	X	X	X	X	X	X	X	?
Self-Concept of Ability	X	X	X	X	X	X	X	?
Children's Self-Concept Scale	X	X	X	X	X	X	X	
IARQ	X	X	X	X	X	X	X	?
Locus of Control Scale for Children	X	X	X	X	X	X	X	
Children's Locus of Control Scale	?	X	X	X	X	X	X	?
Attitude Toward Learning	X	X	X	X	X		X	
Attitude Toward School	X	X	X	X	X		X	
Attitude Toward Reading	?	X	X	X	X	X		X
Quality of School Life	X	X	X	X	X	X	X	?

IV. RECOMMENDATIONS FROM PANEL OF EXPERTS

A panel of experts on measuring affective behavior was convened to make recommendations to SDC. The panel met at SDC on September 10-11, 1975, and consisted of:

Dr. Joyce Epstein, Center for Social Organization of Schools, Johns Hopkins University

Dr. John Kitsuse, Department of Sociology, University of California, Santa Cruz

Dr. Melvin Seeman, Department of Sociology, University of California, Los Angeles

Dr. James Vasquez, Far West Laboratory for Educational Research and Development

During the two-day meeting, the panel discussed issues related to the measurement of affective behavior in the Sustaining Effects Study, examined instruments designed to assess affective behavior, and developed a set of recommendations for the SDC staff. The major recommendations were:

1. Use available instruments rather than develop new ones. The panel felt that while existing instruments all suffered from some shortcomings, they were undoubtedly superior to any that could be developed by SDC in the brief time prior to data collection.

2. Read all instruments to all students, in recognition of the obvious fact that the intent is to assess affective dimensions without contamination by reading ability. While most instruments are intended to be read by the respondent if the respondent has completed third grade, the panel felt that many upper-grade students would lack sufficient skills to handle the reading tasks required; this can be overcome by having instruments read to all students regardless of grade level.

3. Measure affective behavior before measuring cognitive behavior. In considering the total testing schedule for students in the Sustaining Effects Study, the panel strongly urged that affective instruments be administered prior to cognitive instruments (i.e., mathematics and reading achievement tests and measures of functional literacy). The panel felt that students' attitudes would be strongly influenced by immediate occurrences, and that if the affective instruments were administered late in the week of testing, students might express negative views that would not be truly characteristic of them but rather would be in response to the (potentially frustrating) achievement test experiences.

4. Measure students' sense of change over time. The panel suggested that a valuable addition to the assessment of affective behavior would be the use of items that asked the student to indicate the extent to which the student was aware of improvement in his skills and changes in his feelings and attitudes. In proposing this addition, the panel noted that the longitudinal nature of the study included the repeated measurement of students' affective behavior and allowed for the actual consideration of change over time. However, the panel suggested that a particularly useful piece of information would be the student's own awareness of that change. The panel suggested the development of additional items that would indicate the extent to which a student recognized that his skills in reading and mathematics had improved, and that his feelings about himself in the school setting had altered with the passage of time and the experiencing of certain educational activities.

V. INSTRUMENT TO BE USED IN THE STUDY

Following the panel meeting, members of the SDC staff met with personnel from USOE to formalize decisions about measuring affective development. The most critical decision, based on the panel's concern with the adequacy of the recommended instruments, was that the study should focus primarily on the measurement of attitudes toward mathematics and reading. Inasmuch as the

principal objective of compensatory education programs is to improve skills in reading and mathematics, it was judged most appropriate for the Sustaining Effects Study to be especially concerned with students' attitudes toward those two curricular areas. Given this decision, a search was instituted to locate the best existing instrument for attitudes toward mathematics and reading.

The search resulted in the selection of the "Survey of School Attitudes" (Harcourt Brace Jovanovich, 1975) for use in the study. This instrument has been recently developed and released, and is ideally suited to the needs of the Sustaining Effects Study. The "Survey of School Attitudes" (SSA) is designed to measure student reactions to reading and other language arts, mathematics, science, and social studies. Students indicate whether they like, dislike, or are neutral toward different activities in each curricular area. The Survey can be used in group administration settings by a classroom teacher. There are two levels of the Survey: Primary (grades 1-3) and Intermediate (grades 4-8). For the two scales of interest in the Sustaining Effects Study, the following substantive topics are included:

Reading and other language arts: reading, working with words and sounds, writing, speaking, listening.

Mathematics: concepts (of numeration, sets, etc.), computation, geometry and measurement, problem solving, charts and graphs.

The SSA was standardized in 1973 on a sample of 13,500 students in grades one through eight. Twelve school systems in 10 states participated in the standardization. The standardization sample was highly similar to the nation's population, as indicated by 1970 census data, on the following dimensions: geographic region, socioeconomic variables, minority population, and community size.

Item analyses of the mathematics and reading scales yielded the following:

Median Item-Scale Correlations (All Grades Combined)

	Reading	Mathematics
Form A, Reading	.54	.37
Form A, Mathematics	.42	.57
Form B, Reading	.36	.38
Form B, Mathematics	.41	.57

Reading-Mathematics Correlations

	Form A	Form B
Primary Level	.69	.69
Intermediate Level	.43	.44

These data demonstrate that items correlated considerably higher with their own scale than with the other scale, and that there is some degree of commonality in the measurement of attitudes toward reading and towards mathematics.

Reliabilities of the instrument were determined both by Cronbach's Coefficient Alpha procedure and by test-retest with alternate forms over a 10-day interval. Results are summarized below:

Test Form	Alpha Coefficients
Primary Reading, Form A	.81
Primary Reading, Form B	.83
Primary Math, Form A	.85
Primary Math, Form B	.85
Intermediate Reading, Form A	.84
Intermediate Reading, Form B	.82
Intermediate Math, Form A	.92
Intermediate Math, Form B	.90

Scale	Test-Retest Coefficients
Primary Reading	.65
Primary Math	.65
Intermediate Reading	.77
Intermediate Math	.83

These reliability estimates indicate that the SSA yields reasonably stable scores.

Validity of the instrument was approached from several directions. To find out whether the instrument actually measures student achievement, correlations between SSA scores and achievement test scores were computed. They seldom exceeded .30, leading to the conclusion that the SSA measures something different from achievement, and is not overly influenced by achievement. To find out whether students respond in socially-desirable directions, or to please the teacher, scale intercorrelations were inspected to see if they exceeded scale reliabilities (they did not), and teachers were asked whether they felt students responded honestly (they did). These results point to a lack of response bias. Finally, construct validity was approached through a series of factor analyses, which showed the SSA to measure separate attitude dimensions.

To consider minority-group concerns, a substudy was conducted in which minority and non-minority students' scores were compared. The two groups produced comparable results for both reading and mathematics scales, and on both primary and intermediate levels.

Each item of the SSA contains a picture of an activity related to a curricular area, and has an accompanying statement describing the picture. At the Primary level the statements are to be read by the examiner; at the Intermediate level the statements are also printed on the item. The response options are three faces, one smiling, one frowning, and one neutral. By marking the smiling face, the student indicates liking for the activity; marking the frowning face indicates dislike for the activity; and the neutral face means the student is not sure how he or she feels about the activity.

Some modifications of the existing instrument were made for this study. Since science and social studies are not relevant, those scales were eliminated. Scales for reading and mathematics were extended by combining items from the alternate forms of the Survey. The two scales contain 20 items, which should be sufficient to assess changes that take place during the five years of data collection in the study.

In addition, new items were added to the instrument to create scales that measure the following: attitude toward school in general, self-concept in the school setting, and students' sense of change in their own affective behavior over time. These additional items are described in detail later in this paper. The new items, along with the extension of the reading and mathematics subscales, make the length of the total instrument about the same as the original SSA with all four subscales.

The Student Affective Measures for the Sustaining Effects Study have two levels: a Primary version for grades 1-3 and an Intermediate version for grades 4-9. Each version contains 56 items, as follows:

	Number of Items	
	Primary	Intermediate
READING/LANGUAGE ARTS		
Reading	7	5
Working With Words	6	7
Writing	1	2
Speaking	1	2
Listening	2	1
Other Related Activities	3	3
MATHEMATICS		
Concepts of Numeration, Sets, etc.	9	9
Computation	6	4
Geometry and Measurement	4	4
Problem Solving	1	2
Charts and Graphs	0	1
SCHOOL IN GENERAL	4	4
SELF-CONCEPT IN SCHOOL	4	4
CHANGE OVER TIME		
School in General	2	2
Self-Concept	2	2
Mathematics	2	2
Reading	2	2
TOTAL	56	56

The 20 items for the scale on attitude toward mathematics and the 20 items for the reading attitude scale were obtained by using all 15 items from Form A and five items from Form B. Items chosen from Form B are those that bear the least resemblance to items in Form A, to minimize the extent to which students will feel that they have already answered an item. Reading and mathematics items alternate, somewhat reducing the development of response set by students.

To measure attitude toward school in general, four items from the satisfaction-with-school scale of the "Quality of School Life" inventory are used. The four items reflect attitudes that are not tied to subject-matter areas, adults in school, or other students, but rather deal with school and class in a general manner. The items were modified to allow them to fit a response pattern of "Yes," "Not Sure," and "No," using the same three faces that appear in the reading and mathematics scales. Also, items cast in the negative were changed to the positive to eliminate difficulties students have responding to negative items (does one answer "Yes" or "No" to the statement, "I do not like school very much"?). The items are:

I enjoy the work I do in class.
School work is very interesting.
I like school very much.
I am happy when I am at school.

Self-concept in school is assessed by four items selected from existing self-concept scales (Piers-Harris "Children's Self-Concept Scale" and Sears "Self-Concept Inventory") and modified to use the same response pattern ("Yes," "Not Sure," "No"). The particular items chosen refer specifically to the student in the school setting, rather than being related to self-concept in broader contexts such as home or play. The items are:

I can think up answers to questions.
I like to learn about new things.
I am good in my school work.
I can learn things quickly.

The eight items measuring the student's sense of change over time in his affective behavior were prepared following guidelines suggested by the expert panel. They, too, were written to use the response mode described above. The items are:

I like reading more than I used to.
I do better work in reading than I used to.
I like mathematics more than I used to.
I do better work in mathematics than I used to.
School work is more interesting than it used to be.
I like school more than I used to.
I am a better student than I used to be.
I like my school work more than I used to.

A careful examination of existing locus-of-control scales failed to uncover a reasonable set of items that could be included in the affective measures instrument for the Sustaining Effects Study. The better scales (e.g., the Crandall et al. "Intellectual Achievement Responsibility Questionnaire") are cast in the form of an outcome for which the respondent is to choose one of two possible causes. For example: "If something is easy to learn at school, it is because (a) you pay attention, (b) the teacher gives you lots of help." Using items of this type would require a change in the instrument's response format, adding to the time needed for administration. In addition, younger students may experience difficulty in responding accurately to such items. On the other hand, locus-of-control items cast in the "yes-no" format (e.g., "When I do good work in class it is because I am lucky") are not the best indicators of internal-external locus of control.

VI. SPECIAL CONCERNS IN A LONGITUDINAL STUDY

In the Sustaining Effects Study, SDC will address the usual matters in the assessment of affective development, such as validity, reliability, response bias, item discrimination, scale qualities, and administrator capability. But in a longitudinal study there are a number of unique questions requiring attention. Three particular ones have been identified at present.

The first concerns the linearity of affective development. Do students' attitudes develop as a continuous process in essencially equal increments, or is that development characterized by discontinuities with marked plateaus and peaks? The study's twice-a-year testing schedule of the same students for a number of years should provide information relevant to this question, but care must be exercised in analyzing results to guard against drawing conclusions based on fortuitous happenings that are not reflective of the actual development of affective behavior.

A second concern is with the use of repeated measurements with the same instruments. While the scales for attitudes toward reading and mathematics are made up of different items in the primary and intermediate versions they are extremely similar, and identical items are used in both versions for scales of attitude toward school, self-concept, and sense of change in attitudes. It is possible for a student entering the study in the fourth grade and tested for five years to respond to the same items ten times. What will be bred by such an occurrence: contempt, indifference, boredom, frustration? Will sheer familiarity result in students responding in ways that are not representative of their true attitudes? Data on intra-individual responses must be inspected to cast light on this issue, and it is likely that special observations of testing sessions will be carried out to collect anecdotal information on student behavior during administrations of the Student Affective Measures.

The third issue concerns the interrelationships among program, cognitive, and affective variables. The essential questions are, do program characteristics influence simultaneously both cognitive and affective development, does cognitive development influence affective development, or does affective development produce changes in cognitive development? Valid arguments have been offered for all three positions, and a critical analytic issue in this study will be to tease out conclusions related to the interactions among different variables. Should the longitudinal results clearly point to a particular statement about the interdependence of program characteristics, affective changes, and cognitive growth, this statement would have important policy implications for future program specifications.